Co-development of a Best Practice Checklist for Mental Health Data Science: A Delphi Study

Elizabeth J. Kirkham*, Catherine J. Crompton, Matthew H. Iveson, Iona Beange, Andrew M. McIntosh and Sue Fletcher-Watson

Division of Psychiatry, Centre for Clinical Brain Sciences, Kennedy Tower, Royal Edinburgh Hospital, University of Edinburgh, Edinburgh, United Kingdom

Background: Mental health research is commonly affected by difficulties in recruiting and retaining participants, resulting in findings which are based on a sub-sample of those actually living with mental illness. Increasing the use of Big Data for mental health research, especially routinely-collected data, could improve this situation. However, steps to facilitate this must be enacted in collaboration with those who would provide the data - people with mental health conditions.

Methods: We used the Delphi method to create a best practice checklist for mental health data science. Twenty participants with both expertise in data science and personal experience of mental illness worked together over three phases. In Phase 1, participants rated a list of 63 statements and added any statements or topics that were missing. Statements receiving a mean score of 5 or more (out of 7) were retained. These were then combined with the results of a rapid thematic analysis of participants' comments to produce a 14-item draft checklist, with each item split into two components: best practice now and best practice in the future. In Phase 2, participants indicated whether or not each item should remain in the checklist, and items that scored more than 50% endorsement were retained. In Phase 3 participants rated their satisfaction with the final checklist.

Results: The final checklist was made up of 14 "best practice" items, with each item covering best practice now and best practice in the future. At the end of the three phases, 85% of participants were (very) satisfied with the two best practice checklists, with no participants expressing dissatisfaction.

Conclusions: Increased stakeholder involvement is essential at every stage of mental health data science. The checklist produced through this work represents the views of people with experience of mental illness, and it is hoped that it will be used to facilitate trustworthy and innovative research which is inclusive of a wider range of individuals.

Keywords: mental health, data science, health data, participatory research, Delphi, lived experience, co-produced research

INTRODUCTION

Data science, in which knowledge is derived from high volume data sets (1), holds great potential for mental health research (2). Specifically, large quantities of routinely-collected data, such as NHS health records, represent an opportunity to overcome one of the greatest problems previously inherent to such research: recruiting and retaining a representative sample of participants (3, 4). Recruitment in itself is often time consuming and can be one of the most challenging parts of a research study, whilst the recruitment of a representative sample is harder still (5). In mental health research, the representativeness of a sample can be influenced by numerous factors, including clinicians' willingness to refer participants (6), the severity of participants' mental illness (7), and participants' employment status (8). Of particular concern is evidence that people from ethnic minorities are under-represented in mental health research (9, 10), despite having higher rates of diagnosis for some conditions, such as psychosis (11). These factors mean that the groups of individuals who take part in mental health research studies rarely represent the population of people living with mental illness (12), which could have serious implications for treatment outcomes (13).

Increased use of routinely-collected mental health data is likely to make research more inclusive and to contribute to the development of more tailored treatments (1, 13). However, work which uses routinely-collected health data, especially sensitive mental health data (14, 15), relies upon the trust of the public whose data are being accessed (16) – by definition, analysis of routinely-collected data does not involve informed consent from the individuals who provide such data. This means that researchers who work with mental health data must understand and incorporate the views of people with experience of mental illness in their research practice; after all, it is these individuals whose lives mental health research seeks to improve (17). Such consultation is especially timely in light of the ongoing rapid expansion within mental health data science.

To this end, we sought to generate a best practice checklist for use in mental health data science to support research that is both rigorous and trusted by those who provide mental health data. The checklist was designed to complement other guidance regarding good practice within data science, such as the UK Data Ethics Framework (18), the UK Government's Code of Conduct for Data-Driven Health and Care Technology (19), and recent work on the development of data governance for the use of clinical free-text data (20). Its unique contribution is to encapsulate the perspective of people with lived experience of mental illness, without making recommendations that contravene existing data science frameworks.

To create a checklist that enshrines the principles of trustworthiness and patient-driven priorities within existing research practice, we worked directly with people with expertise in both mental illness and data science. We chose to use the Delphi method, an iterative process in which a group of experts anonymously contributes to the development of consensus on a given topic (21). Delphi studies, which typically recruit between 15 and 30 experts (22), have previously been used to derive guidelines in mental health-related areas such as post-disaster

psychosocial care, and first aid recommendations for psychosis and suicidal ideation (23–26). Whilst some previous studies have sought to consolidate the views of distinct groups of stakeholders (27), we took the approach of recruiting people with personal or professional expertise in both mental illness and data science. This ensured that the participants themselves were in a position to weigh up the relative merits of the information from both perspectives, and reduced the need for researcher involvement in handling potential trade-offs. The Delphi took place over three phases; each phase involved an online survey completed by participants, followed by analysis by the research team and creation of the next survey (**Figure 1**).

METHODS

## Pre-design Phase

A pre-design phase was used to identify topics, and later, statements, to be included in the first phase of the Delphi study. First the research team presented a broad outline of the purpose of the research to a group of stakeholders (four people with lived and professional experience of mental illness and two psychiatry researchers), and asked them to write down suggestions for what was important for best practice in mental health data science. These stakeholders were independent of the participants who took part in the three key phases of the Delphi study. The stakeholders were asked to write ideas on "Post-It" notes, and stick them onto four large sheets of paper titled "Concerns," "Hopes," "Best practice," and "Other ideas." The resulting ideas covered concepts such as risk management, data sharing, and agency of participants.

After this meeting, the research team drew upon the stakeholders' ideas and their own experiences working in mental health and data science research to generate statements covering the topic of best practice in mental health data science. This led to the creation of 73 statements which loosely fell into the following categories: users of data, access to data, data linkage, anonymity and de-identification, consent, governance, and community. The statements were compiled into a draft version of the Phase 1 survey and presented to the stakeholder group discussed above. The stakeholders recommended improvements primarily based around clarity and usability, and the research team used these to create the final version of the Phase 1 survey.

Participants

Participants were recruited if they had both experience with mental illness and experience with data science or research methods. The criteria used to assess this experience, and the number of participants within the final sample who selected each criterion, are listed in **Table 1**. All participants in the final sample had personal experience of mental illness, and 80% of participants indicated that they were living with a mental illness at the time of the study. Before entry into the study potential participants were asked by email to confirm that they met at least one of the criteria in each category. Nineteen of the 20 participants in the final sample fulfilled more than one of the criteria for inclusion in terms of their experience with mental illness, and all participants

FIGURE 1 | Illustration of the phases of the Delphi process.
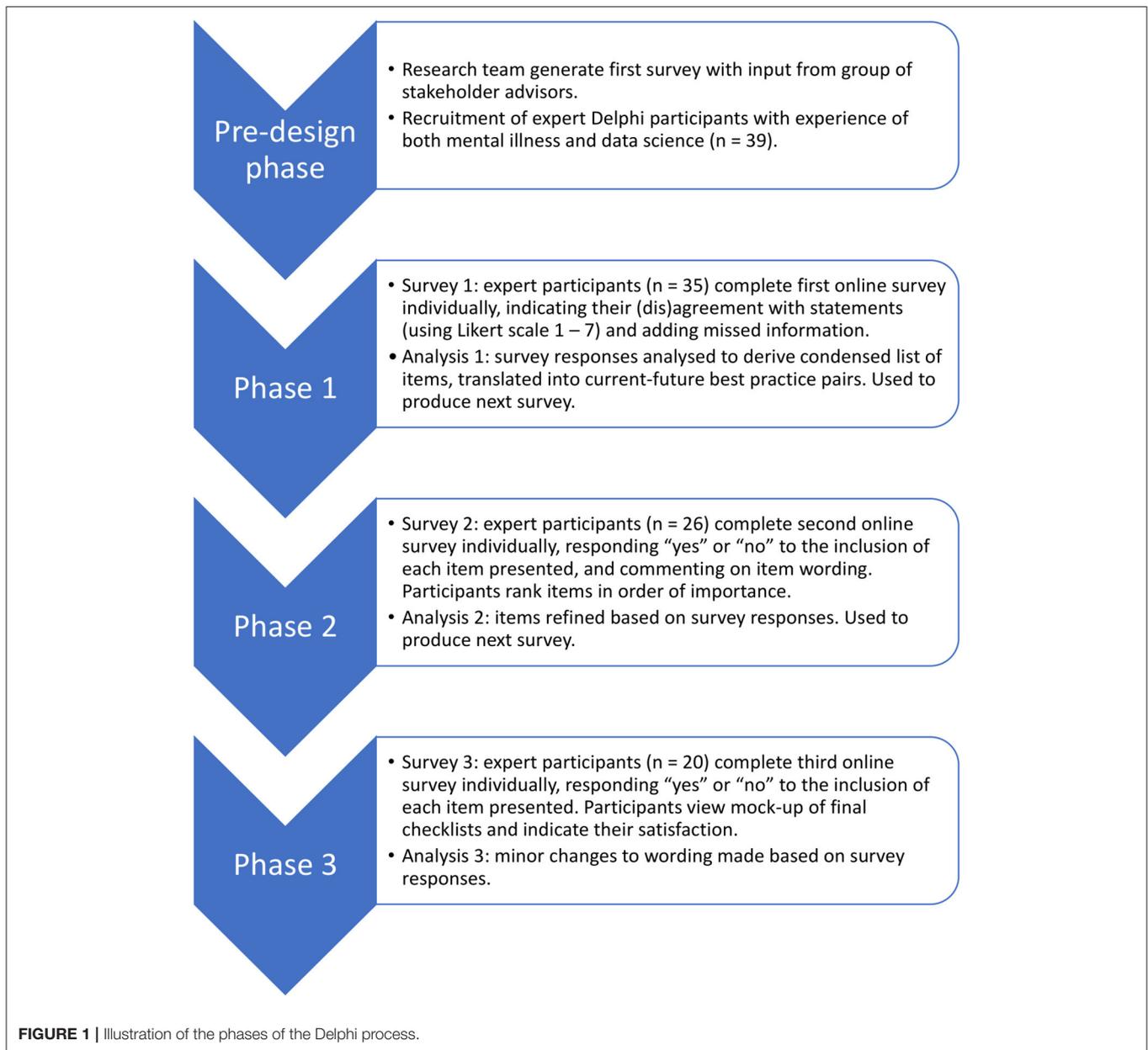
fulfilled more than one of the criteria for inclusion in terms of their experience with data science or research methods.

Participants were recruited through a snowballing technique with relevant contacts and through social media. Thirty-nine participants agreed by email to take part in the study. Of these, 35 individuals provided sufficient data for analysis, though one participant was excluded due to insufficient experience of mental illness and four participants were excluded due to a technical error. Consequently, 30 participants were included in Phase 1 of the study. Of these, 26 provided sufficient data for analysis during Phase 2 of the study, and of these, 20 provided data for Phase 3. Participants who did not provide data for a given phase were excluded from subsequent phases. It should be noted that data collection coincided with the start of the Covid-19

pandemic and UK lock-down, which could have increased drop-out. Nevertheless, a sample of 20 participants is considered a good sample size for a Delphi study (22).

Overview of Phases

An illustrative overview of the phases included in the Delphi study can be found in **Figure 1**. The research received ethical approval from the School of Health in Social Science Ethics Committee, University of Edinburgh, ref STAFF212. Potential participants were sent an overview information sheet by email, and those who chose to take part were provided with this overview again at the start of the first survey. All participants provided informed consent to the whole Delphi study at the beginning of the first online survey by responding by tick box

TABLE 1 | Participants' experience relevant to mental illness and data science or research methods.

Experience relevant to mental illness	Number	Experience relevant to data science or research methods	Number
I have been diagnosed with a mental illness at some time in my life.	20	I have an undergraduate qualification in an area relevant to data science or research methods (e.g., psychology, clinical science, epidemiology, statistics etc.).	13
I consider myself to have a mental illness at the moment.	16	I have a postgraduate qualification in an area relevant to data science or research methods (e.g., psychology, clinical science, epidemiology, statistics etc.).	14
I have family members or close friends who live with mental illness.	15	I have advised on a research study/worked on a research team.	17
I work in an occupation which is related to mental health.	10	I work or have worked as a peer researcher.	6
I am a mental health practitioner.	2	I work or have worked in a research setting.	15
		I work or have worked in a setting that interfaces with research (e.g., third sector organization, NHS).	14

Participants included in this table are those who took part in all three phases, n = 20. All participants fulfilled at least one criterion per column (mental illness and data science). Both columns total more than 20 because 95% of participants fulfilled multiple criteria in the mental illness column, and all participants fulfilled multiple criteria in the data science and research methods column.

TABLE 2 | Terminology used to describe survey contents.

Term	Use in text	Phase
Statement	Refers to one of the 63 statements included in the Phase 1 survey	1
Category	Refers to the seven categories used to sort the 63 statements in Phase 1	1
Sub-category	Refers to the sub-sections of the seven categories used in Phase 1	1
Item	One of the 14 items included in the (draft) checklist in Phases 2 and 3. Each item is divided into two components, a "best practice now" component and a "best practice in the future" component.	2 and 3
Component	Each of the 14 checklist items is divided into two components, a "best practice now" component and a "best practice in the future" component. Therefore, there are 28 components in total within the best practice checklist.	2 and 3

to a series of consent statements. An additional, phase-specific information sheet was provided at the beginning of each of the three phases. All data were collected online using surveys hosted by Qualtrics (www.qualtrics.com). In all three surveys, words whose meaning could have been unfamiliar or ambiguous were highlighted in red, and defined in a glossary that was available to download on each page of the survey (**Appendix 1**).

Phase 1
Materials and Procedure
Table 2 provides definitions for the terms used in the subsequent paragraphs. Participants were given 10 days to complete the first survey, and were sent reminder emails during this period. The survey presented participants with a series of 63 statements (**Appendix 2**). The statements were organized into seven categories (users of data, access to data, data linkage, anonymity and de-identification, consent, governance and community), each of which was divided into two or three sub-categories (**Table 3**). Each sub-category contained between 2 and 5 statements, and each category (i.e., the combination of sub-categories) contained between 8 and 10 statements. Each category was presented on a separate page of the survey, resulting in between 8 and 10 statements per page. Participants were asked to indicate the extent to which they believed that the given statement represented best practice for mental health data science, using a 7-point Likert scale which ranged from

"strongly disagree (1)" to "strongly agree (7)." At the end of each category of statements, participants were presented with a text box in which they could make comments on the wording of the statements within that category. At the end of the survey, participants were presented with an additional text box in which they could enter any topics or statements concerning best practice in mental health data science that had been missed during the survey.

Analysis Methods
We began Phase 1 with a large quantity of statements ($n = 63$, see **Appendix 2**), some of which contradicted one another, in order to cover as many potential viewpoints as possible. As described above, the original statements were organized by category and sub-category (**Table 3**). The initial aim of the Phase 1 analysis was to prune and condense the extensive list of statements into one statement per sub-category. The Phase 1 analysis took place over two stages, which we refer to as Stage A and Stage B below (note these "stages" are distinct from the "Phases," the latter of which represent timeframes of data collection and analysis).

Stage A
Stage A involved the initial processing of the quantitative (participant scoring of statements) and qualitative data (participant comments). It was decided prior to analysis that statements which received clear support would be retained for Stage B of Phase 1 analysis. "Clear support" in this instance was

TABLE 3 | Categories and sub-categories for statements and items included in the surveys.

Category	Survey sub-categories (Phase 1)	Survey sub-categories (Phases 2 and 3)
1. Users of data	Who uses data	Who uses data
	Where data are accessed	Where data are accessed
	Checks on data users/how access is monitored	Checks on data users/how access is monitored
2. Access to data	Giving access	Giving access
	Getting access	Getting access
	Use of synthetic data	
3. Data linkage	*Linking mental health data with data from other public services*	**Data linkage**
	Linking mental health data with individually-created data	
4. Anonymity and de-identification	De-identifying data	De-identifying data
	Protecting against accidental identification	Protecting against accidental identification
5. Consent	*Giving people control over their data*	**Consent**
	Ensuring maximum access for scientific purposes	
	Using alternative models of consent	
6. Governance	Dealing with requests for data withdrawal	Dealing with requests for data withdrawal
	How we respond to mistakes	How we respond to mistakes
	How we enact quality control	How we enact quality control
7. Community	Ensuring public trust in mental health data science	Ensuring public trust in mental health data science
	How we understand the context in which mental health data science occurs	How we understand the context in which mental health data science occurs

Sub-categories that were removed are highlighted in italics, sub-categories that were added are highlighted in bold. The other sub-categories remained the same throughout.

defined as a mean score of more than 5 out of 7, indicating a level of agreement higher than "somewhat agree." Statements which received clear disagreement (mean score of <3, lower than "somewhat disagree") would be either removed or, alternatively, retained for Stage B but reversed so that they represented the opposite position (28). Statements with a mean score ≥3 and ≤5 would be discarded, as these scores represented opinions which were clustered around "neutral," suggesting the absence of a consensus opinion from participants. We defined the aforementioned numerical thresholds prior to data collection, and chose to use the mean rather than the median in order to allow for more nuanced distinctions when comparing statements' average ratings. Next, after quantitative responses to the statements had been analyzed, we continued with Stage A by processing the qualitative text responses provided by participants. These data were analyzed using a "rapid assessment" version of thematic analysis (29), in which the content of each response was coded and common themes were extracted.

Throughout Stage A analysis, the data were processed within the sub-categories that were present in the Phase 1 survey (**Table 3**). At the end of Stage A, it was apparent that some of these sub-categories were not useful. For example, quantitative and qualitative responses to the three "consent" sub-categories used in the Phase 1 survey (individual control, scientific access, and alternative models; **Table 3**) indicated that participants' views of consent did not fit into these three sub-categories. As a result, a single "consent" category was used instead. In this manner, where appropriate, the sub-categories used in Phase 1 were updated (**Table 3**).

Once Stage A was complete, we were left with a dataset divided into 14 sub-categories. Each sub-category contained a list of statements that had received clear support (mean score of more than 5 out of 7), and additional recommendations that had been derived from participants' qualitative responses. Taking for example the sub-category "where data are accessed," one of the three statements was retained ("...only allowing researchers to see data in specific digital environments (a.k.a. 'safe havens')"; see **Appendix 2** for statements that were discarded). In addition, the data for this sub-category contained the recommendation, derived from participants' qualitative responses, that digital controls were preferable to physical controls, a comparison that had not been explicitly addressed within the statements themselves.

Stage B
In Stage B of the Phase 1 analysis, all the data for a given sub-category (retained statements and additional information, as described above) were collated. The initial aim of Stage B was to distill all the data within each sub-category into one checklist item. However, when examining the data, it became apparent that in many cases the data collated for each sub-category contained two distinct elements of best practice: recommendations that could be implemented immediately, and those that would depend upon future development and supporting infrastructure. Therefore, instead of following the initial aim of creating one overarching checklist item from the data contained within each sub-category, we decided to create an item which would contain two components: one that referred to best practice that could be implemented immediately, and one that referred to "ideal"

best practice to be put in place in future. For example, for the sub-category "responding to mistakes," the best practice item was divided into a current component: "Best practice for mental health data science means planning in advance to avoid data breaches, utilizing a recording process for data breaches, and reporting near misses," and a future component: "Best practice for mental health data science means developing robust systems to prevent data leaks and breaches." The newly-created checklist items for each sub-category were then used to create the Phase 2 survey, described below.

Phase 2

Materials and Procedure

Participants were given 11 days to complete the second survey, and were sent reminder emails during this period. Participants were presented with the new items (each with a "now" and a "future" component) for each of the 14 sub-categories. For example, one "now" component was "Best practice for mental health data science means allowing other researchers to check analyses wherever possible." Its corresponding "future" component was "Best practice for mental health data science means providing access to synthetic data where real data cannot be shared, in order to allow other researchers to check analyses." The full list of items included in this second survey is illustrated in **Table 4**.

For each component, participants were asked to respond "yes" or "no" to whether it should be included in the relevant checklist. After each statement, participants were provided with a text box in which they could make comments on the wording of the component. After this, participants were presented with all the components relating to current best practice (the "now" checklist) and asked to organize them in order of importance. This was then repeated with the future best practice components (the "future" checklist). These rankings are presented in **Figures 2**, **3**.

Analysis Methods

The data from the second survey were analyzed using both quantitative and qualitative methods. The data used for the quantitative analysis were the yes/no responses provided for each component. Based on decisions made prior to analysis, components which received a "yes" response from fewer than 50% of participants were discarded, as this indicated that less than half of the group wanted the statement to be included. The qualitative responses, entered in the aforementioned text boxes, were classified into three categories: meaningful recommendation, formatting/language comment, not relevant/refers to different statement. The content of these comments was then used to refine the existing components, and the updated list of components was used to produce the Phase 3 survey.

Phase 3

Materials and Procedure

Participants were given 8 days to complete the third survey, and were sent reminder emails during this period. Participants were presented with the updated items derived during the analysis stage of Phase 2. As before, two components (corresponding to current and future best practice) were presented for each of the 14 items. Participants were asked if the given component should be included in the relevant final best practice checklist, and could choose "yes," "no," or "other." If a participant chose "other," they were presented with a text box in which they could indicate how the component should be changed. Participants were then asked to open a pdf file of a mock-up version of the current best practice checklist, and asked to indicate how satisfied they would be (on a 5-point scale from "very dissatisfied" to "very satisfied") with a professionally-designed version of this mock-up. This process was repeated for the future best practice checklist.

Analysis Methods

Based on a threshold chosen prior to analysis, statements which received a "yes" response (as opposed to "no" or "other") from fewer than 50% of participants were discarded, as this indicated that less than half of the group wanted the statement to be included. In addition, participants' comments were assessed and minor changes to the wording of the statements were made. Finally, the median level of satisfaction with each mock-up checklist was measured.

RESULTS

Phase 1

Stage A

As described in the Methods, the first stage of Phase 1 analysis, Stage A, involved processing the quantitative and qualitative data collected during the Phase 1 survey. Of the 63 original statements, 12 had a mean value between 3 and 5 and were therefore discarded (**Appendix 2**). The remaining 51 original statements had a mean value greater than 5 and were retained (**Appendix 2**). None of the original statements received a mean score of <3. Next, participants' comments were labeled and sorted into themes by EJK, following training from SFW. These themes provided a way of extracting general recommendations from the set of individual participant responses.

Stage B

The data (included statements and participant recommendations) were grouped by sub-category. Whilst the initial aim had been to produce one checklist item for each sub-category, it was decided that two components, a "current" and a "future" statement, would be created for each item (see Methods for details). These components are presented in **Table 4**.

Phase 2

The list of components included in the Phase 2 survey can be seen in **Table 4**. All components were given a mean "yes" rating higher than 50%; with the lowest rating at 56%, the highest at 100% and the majority (23/28) falling at or above 75% (see **Appendix 3**). Therefore, no components were discarded. The number of text responses to each component is presented in **Table 4**, organized by classification (meaningful recommendation, formatting/language, not relevant/refers to a different statement). These comments were used to update the

TABLE 4 | Phase 2 survey components with their corresponding number of participant comments, divided by comment classification.

| Current or future	Best practice for mental health data science means…	Meaningful recommendation	Formatting/language comment	Not relevant/ refers to different statement
Current	Best practice for mental health data science means data should be accessible to a range of people who conduct research, including academics and health workers.	5	2	0
Future	Best practice for mental health data science means providing appropriate training and supervision for data users, and carrying out criminal record checks.	4	2	0
Current	Best practice for mental health data science means ensuring that data are accessed in safe settings, but that procedures should not be too complicated (to avoid encouragement of unsafe "workarounds").	3	5	0
Future	Best practice for mental health data science means providing digital controls to allow remote access from private settings, using procedures that are not too complicated.	2	5	0
Current	Best practice for mental health data science means creating data management plans and ensuring that these are adhered to over time.	1	6	0
Future	Best practice for mental health data science means incorporating inspection processes to ensure ongoing compliance with good data practice, and responding proportionately to inappropriate data use with measures such as temporary or long-term suspension of access.	4	3	0
Current	Best practice for mental health data science means researchers, scientists and clinical services making data and findings (including null results) open-access where possible, but taking extra care when making decisions regarding access to qualitative data such as free text information.	6	5	1
Future	Best practice for mental health data science means building plans for data collected by researchers, scientists and clinical services to be made available for analysis on an open-access basis.	6	1	0
Current	Best practice for mental health data science means allowing other researchers to check analyses wherever possible.	5	0	0
Future	Best practice for mental health data science means providing access to synthetic data where real data cannot be shared, in order to allow other researchers to check analyses.	5	1	0
Current	Best practice for mental health data science means responsible linking of mental health data with other sources of public data, such as education or welfare data, in order to provide new information of public benefit about mental health.	4	2	0
Future	Best practice for mental health data science means developing effective measures, including secure linking systems, to protect against identification and misuse.	3	1	0
Current	Best practice for mental health data science means using de-identified data, except where identifiable information (including information about protected characteristics) is essential to beneficial outcomes. In all cases the health and benefit of people with lived experience should be prioritized.	2	4	0
Future	Best practice for mental health data science means developing methods for de-identification, including innovative ways to mask identifiable information.	2	1	0
Current	Best practice for mental health data science means incorporating rules-based statistical disclosure control.	3	7	1
Future	Best practice for mental health data science means incorporating principles-based statistical disclosure control with training and external oversight.	4	4	1
Current	Best practice for mental health data science means ensuring that participants have as much control over consent as possible.	4	1	0
Future	Best practice for mental health data science means exploring alternative models of consent, such as blanket consent for a research topic (e.g., drug development for depression), or blanket consent for a type of data being accessed (e.g., blood test data).	8	1	0
Current	Best practice for mental health data science means ensuring that researchers have a process in place for responding to withdrawal requests and that they provide transparency on whether, how and when participants can withdraw their data.	0	0	0

(Continued)

TABLE 4 | Continued

Current or future	Best practice for mental health data science means...	Meaningful recommendation	Formatting/language comment	Not relevant/ refers to different statement
Future	Best practice for mental health data science means appointing an independent arbiter to arbitrate on complex questions relating to consent and data withdrawal.	5	0	0
Current	Best practice for mental health data science means planning in advance to avoid data breaches, utilizing a recording process for data breaches, and reporting near misses.	1	2	0
Future	Best practice for mental health data science means developing robust systems to prevent data leaks and breaches.	5	1	0
Current	Best practice for mental health data science means monitoring data quality and taking account of the origin and quality of data when drawing conclusions.	1	0	0
Future	Best practice for mental health data science means incorporating both stakeholder and procedural oversight of data repositories, with the latter tasked with monitoring data quality and responding to public questions.	5	5	0
Current	Best practice for mental health data science means incorporating the views of people with lived experience throughout the course of each project, and providing sensitive and high quality public communication of findings.	2	2	0
Future	Best practice for mental health data science means following the principles of open access throughout; publicly pre-registering studies, providing online information of each overarching request to use data and consequent outputs, and publication of null results.	4	3	1
Current	Best practice for mental health data science means ensuring that data users understand the underlying data collection tools as well as the socio-cultural context in which studies are designed and findings are disseminated.	4	0	0
Future	Best practice for mental health data science means active commitment and working to reduce stigma associated with the phenomena being studied and to increase public understanding of science.	8	1	0

Number of participants was 26.

FIGURE 2 | Mean ranking of statements in the current best practice checklist during Phase 2, ordered by median ranking. Lower scores (at the top of the figure) indicate higher importance (number of participants = 25).

FIGURE 3 | Mean ranking of statements in the future best practice checklist during Phase 2, ordered by median ranking. Lower scores (at the top of the figure) indicate higher importance (number of participants = 24).
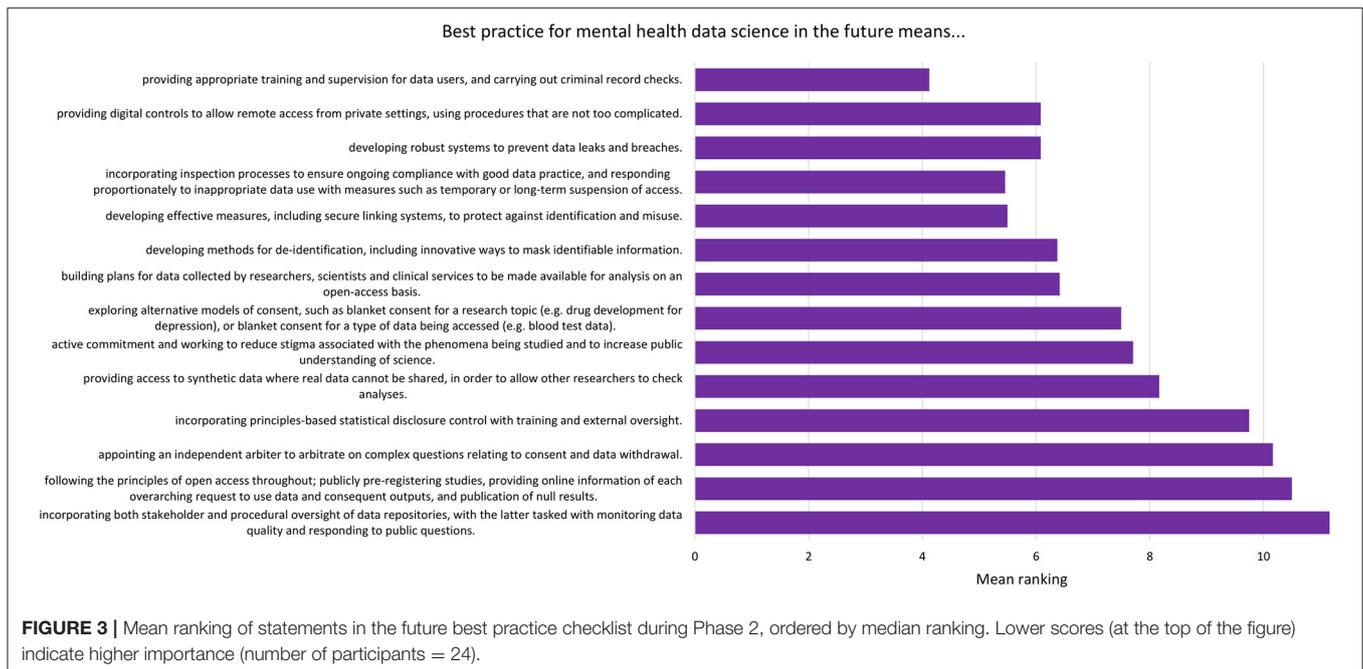
components for use in Phase 3. Mean and median rankings of component importance were calculated for the "right now" and "in the future" checklists (**Figures 2**, **3**). Finally, satisfaction ratings for the two checklists were observed (**Figure 4**). The median satisfaction rating for both checklists was "satisfied."

Phase 3

All statements were given a mean "yes" rating higher than 50%; with the lowest rating at 65%, the highest at 100% and the majority (24/28) falling at or above 75% (see **Appendix 4**). Therefore, no statements were discarded. Minor comments regarding the wording of the statements were considered and the statements were altered where appropriate. Satisfaction ratings for the two checklists are illustrated in **Figure 5**. The median satisfaction rating for both checklists was "satisfied." The final versions of the two checklists can be viewed at https://osf.io/9u8ad/.
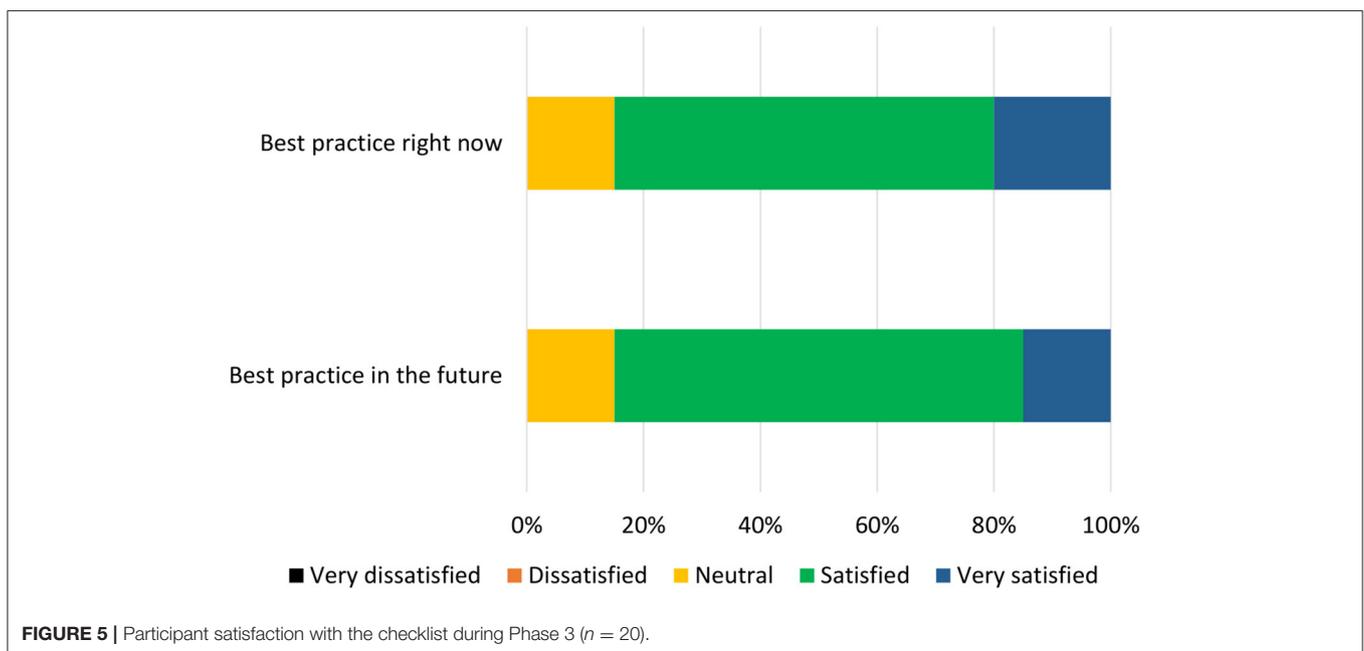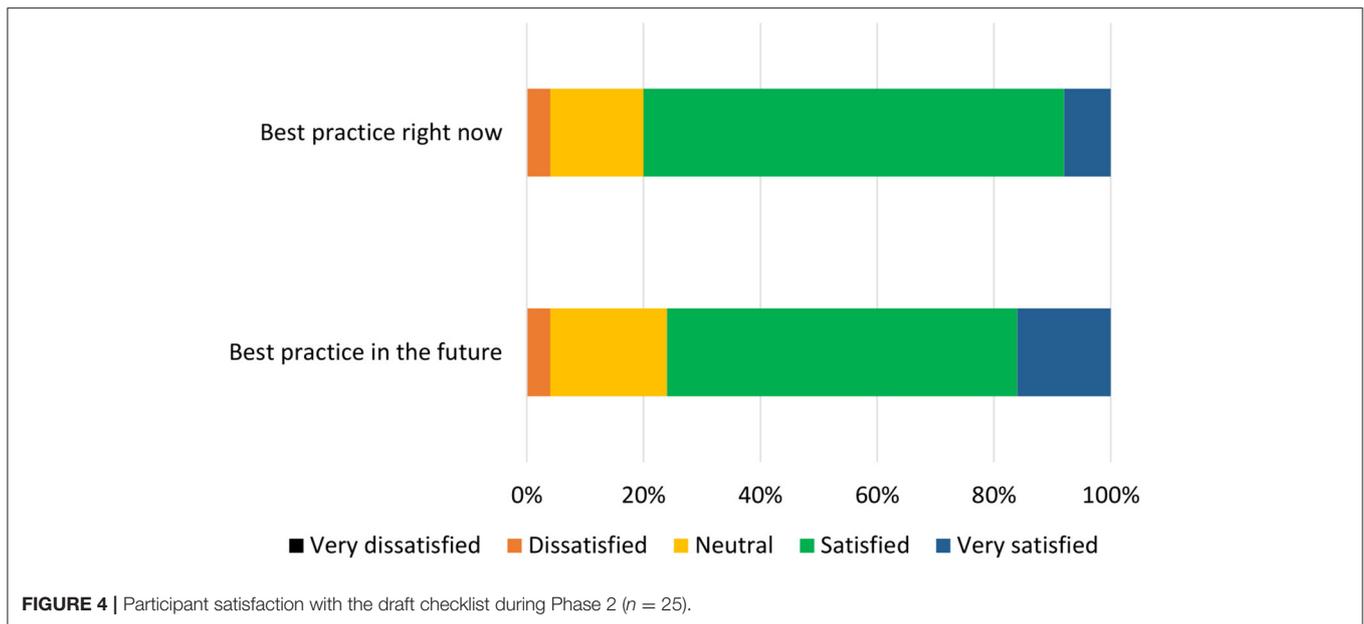
DISCUSSION

This research used the Delphi method to create guidelines for best practice in mental health data science. A group of participants with expertise in mental illness and data science contributed responses over three phases to produce two checklists: one focused on what mental health researchers can do now, and one focused on what the wider mental health data science community can put in place in the future. Each checklist features 14 items that cover issues pertaining to de-identification of data, data security, transparency and oversight, and community perspectives. Each final checklist was approved by all remaining participants, receiving comprehensive support at the level of individual statements and for the checklist as a whole.

As mental health data science moves forward, it is essential to obtain and maintain the trust of those who provide the data, and to represent their views in how data science is conducted (17). The use of the Delphi methodology ensures that participants' involvement goes beyond consultation and becomes part of the scientific process itself. We anticipate that the resultant checklists will be used in conjunction with other work which seeks to apply high standards of data governance to the rapidly evolving field of data science, such as Jones et al.'s (20) position paper on standards for the use of clinical free-text data.

Examination of the rankings for the current best practice checklist (**Figure 2**) showed that de-identification of data was expert participants' highest priority for researchers currently working with mental health data. It is possible that this could reflect particular concerns regarding privacy of mental health data (14), though previous research with the general public suggests that de-identification [which is sometimes perceived as anonymisation; (16)] is also important when considering health data more generally (16, 30, 31). The rankings also demonstrated an emphasis on keeping data safe and secure, though making data accessible was also viewed as important. This pattern of responses complements previous research with the general public, who were supportive of health data sharing assuming certain conditions (such as trust in those handling the data) were met (16).

The top three ranked statements for the future best practice checklist all fell under the sub-category "users of data," covering who accesses the data, where it is accessed, and how this access is monitored. The top three statements (**Figure 3**) reflect a desire for a landscape in which data are well-protected via training and supervision for those using the data, and via ongoing oversight of projects. There was a sense that sanctions for inappropriate use of data should be used, but that these should be in proportion to

FIGURE 4 | Participant satisfaction with the draft checklist during Phase 2 ($n = 25$).

FIGURE 5 | Participant satisfaction with the checklist during Phase 3 ($n = 20$).

the "offense"; for example, an innocent mistake should be treated differently to deliberate misuse of data, and whole research groups should not be sanctioned based on the behavior of one individual.

Whilst data protection was clearly of central importance to the expert participants, it became apparent that some current practices for accessing data can be cumbersome and inefficient, especially where physical procedures, such as attending safe settings, are required before access is gained. As one participant noted, this could unintentionally introduce barriers for certain groups of people, such as those with access needs, who may

require access to the data. Furthermore, there was concern that extensive and inconvenient barriers to data access could encourage unsafe workarounds, such as using the same password for multiple situations. Delphi participants were generally in favor of moving away from physical controls and toward more efficient, digital control of data. The innovation of more secure, streamlined access to mental health data would have wide-ranging benefits, not least within academia where the time it takes to access data can out-run the length of a project that seeks to study such data (17, 32). This is an important factor for the mental health data science

community to consider when designing future procedures and systems.

With respect to the topic of consent, although all participants had expertise in data science, none of the Phase 1 statements concerned with consent procedures which maximize access for scientific research received sufficient consensus for inclusion in later phases (**Appendix 2**). Consent statements which were retained tended to favor giving individuals control over their data, though a number of individual participants recognized the inherent difficulties with this approach (such as restrictions on research and excessive burden on individuals). Dealing with consent is arguably one of the most challenging issues for the mental health data science community, though innovative models are being developed (33, 34). With respect to our research, the expert participants as a group rejected the concept of allowing consent to be provided by a representative sample of participants rather than by all participants, but endorsed the possibility of moving away from individualized models of consent in the future.

With respect to limitations, we acknowledge that there was some drop-out across the three phases, due both to technical error (four participants) and participant attrition (11 participants from Phase 1 to Phase 3). It could be suggested that some of the participants left the study due to disagreement with the concept behind the study. However, this is unlikely given that participants were asked to review an information sheet describing the aims and procedures sent to them by email before they agreed to begin the study. It is also possible that some of the participant attrition was connected to participants' mental health, which could have left a final sample who were more well than the sample that initially agreed to take part. Nevertheless, the vast majority (80%) of the final sample considered themselves to have a mental illness at the time of the research, suggesting that the views of people with current lived experience were represented. Similarly, we cannot be sure of the extent to which the participants in the present study were representative of people with experience of mental illness more generally; by definition—due to the data science experience criterion—our sample were more highly educated than the general population, and although participants were able to provide additional information about their mental health if they chose to, we did not systematically collect data on the specific mental health conditions they experienced. It is therefore possible that the sample did not represent the full range of people with mental health conditions whose data may be included in future mental health data science research. In addition, the useability of the checklist would benefit from feedback from experts in countries outside the UK. Finally, whilst a key strength of our study was the inclusion of expert participants with both experience of mental illness and professional knowledge of data science, it is possible that this group of people may differ in their views from those with experience of mental illness but less knowledge of data science. Having said this, given that our study was designed to create guidelines for mental health data science rather than gather opinions on the topic, this is perhaps a less pressing concern.

The underuse of increasingly large sources of data is arguably leading to avoidable health harms (35), not least in mental health research, where big data are less widely used than in fields such as oncology or cardiology (1). However, as mental health data science develops, it is essential that those with experience of mental illness are included every step of the way. Future work may seek to employ co-design workshops or other participatory activities to do this; here, we chose to use a participatory Delphi method. Our two resultant checklists focus, respectively, on what mental health researchers working now can do to make their data trustworthy, and on the actions the wider mental health data science community should take in the future. By "in the future," we refer especially to new platforms for mental health data science; such developments should aim to adhere to the advice provided by the future checklist, and to use it as a guide in the creation of new infrastructure. The rapidly-growing opportunities for using routinely-collected mental health data offer the chance for more inclusive research which captures information from those who are, for whatever reason, unable to engage with traditional research methodologies (4, 7). We hope that these checklists will facilitate such use, in turn supporting the development of outcomes which include and benefit those who need them most.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by School of Health in Social Science Research Ethics Committee at the University of Edinburgh. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

SF-W, AM, and CC contributed to the conception of the study. EK, SF-W, IB, CC, and MI contributed to the design of the study. EK and SF-W performed the analysis. EK wrote the first draft of the manuscript. CC, SF-W, IB, and MI made revisions to the draft manuscript. All authors read and approved the submitted version.

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyt.2021.643914/full#supplementary-material

REFERENCES

1. McIntosh AM, Stewart R, John A, Smith DJ, Davis K, Sudlow C, et al. Data science for mental health: a UK perspective on a global challenge. *Lancet Psychiatry.* (2016) 3:993–8. doi: 10.1016/S2215-0366(16)30089-X

2. Simon GE. Big data from health records in mental health care hardly clairvoyant but already useful. *JAMA Psychiatry.* (2019) 76:349–50. doi: 10.1001/jamapsychiatry.2018.4510

3. Furimsky I, Cheung AH, Dewa CS, Zipursky RB. Strategies to enhance patient recruitment and retention in research involving patients with a first episode of mental illness. *Contemp Clin Trials.* (2008) 29:862–6. doi: 10.1016/j.cct.2008.07.005

4. Woodall A, Morgan C, Sloan C, Howard L. Barriers to participation in mental health research: are there specific gender, ethnicity and age related barriers? *BMC Psychiatry.* (2010) 10:10. doi: 10.1186/1471-244X-10-103

5. Martin AM, Zakrzewski JJ, Chou, CY, Uhm SY, Gause RM, et al. Recruiting under-represented populations into psychiatric research: results from the help for hoarding study. *Contemp Clin Trials Commun.* (2018) 12:169–75. doi: 10.1016/j.conctc.2018.11.003

6. Patterson S, Kramo K, Soteriou T, Crawford MJ. The great divide: a qualitative investigation of factors influencing researcher access to potential randomised controlled trial participants in mental health settings. *J Mental Health.* (2010) 19:532–41. doi: 10.3109/09638237.2010.520367

7. Lally J, Watkins R, Nash S, Shetty H, Gardner-Sood P, Smith S, et al. The representativeness of participants with severe mental illness in a psychosocial clinical trial. *Front Psychiatry.* (2018) 9:7. doi: 10.3389/fpsyt.2018.00654

8. Kannisto KA, Korhonen J, Adams CE, Koivunen MH, Vahlberg T, Välimäki MA. Factors associated with dropout during recruitment and follow-up periods of a mhealth-based randomized controlled trial for mobile.net to encourage treatment adherence for people with serious mental health problems. *J Med Internet Res.* (2017) 19:e46. doi: 10.2196/jmir.6417

9. Gulsuner S, Stein DJ, Susser ES, Sibeko G, Pretorius A, Walsh T, et al. M. Genetics of schizophrenia in the South African Xhosa. *Science.* (2020) 367:569–73. doi: 10.1126/science.aay8833

10. Iwamasa GY, Sorocco KH, Koonce DA. Ethnicity and clinical psychology: a content analysis of the literature. *Clin Psychol Rev.* (2002) 22:931–44. doi: 10.1016/S0272-7358(02)00147-2

11. Coid JW, Kirkbride JB, Barker D, Cowden F, Stamps R, Yang M, et al. Raised incidence rates of all psychoses among migrant groups: findings from the East London first episode psychosis study. *Arch Gen Psychiatry.* (2008) 65:1250–8. doi: 10.1001/archpsyc.65.11.1250

12. Kline E, Hendel V, Friedman-Yakoobian M, Mesholam-Gately RI, Findeisen A, Zimmet S, et al. A comparison of neurocognition and functioning in first episode psychosis populations: do research samples reflect the real world? *Soc Psychiatry Psychiatr Epidemiol.* (2019) 54:291–301. doi: 10.1007/s00127-018-1631-x

13. Iltis AS, Misra S, Dunn LB, Brown GK, Campbell A, Earll SA, et al. Addressing risks to advance mental health research. *JAMA Psychiatry.* (2013) 70:1363–71. doi: 10.1001/jamapsychiatry.2013.2105

14. King T, Brankovic L, Gillard P. Perspectives of Australian adults about protecting the privacy of their health information in statistical databases. *Int J Med atics.* (2012) 81:279–89. doi: 10.1016/j.ijmedinf.2012.01.005

15. Martínez C, Farhan I. *Making the Right Choices: Using Data-Driven Technology to Transform Mental Healthcare.* Westminster, CO: Reform, (2019).

16. Aitken M, Jorre JD, Pagliari C, Jepson R, Cunningham-Burley S. Public responses to the sharing and linkage of health data for research purposes: a systematic review and thematic synthesis of qualitative studies. *BMC Med Ethics.* (2016) 17:73. doi: 10.1186/s12910-016-0153-x

17. Ford E, Boyd A, Bowles JKF, Havard A, Aldridge RW, Curcin V, et al. Our data, our society, our health: a vision for inclusive and transparent health data science in the United Kingdom and beyond. *Learning Health Syst.* (2019) 3:12. doi: 10.1002/lrh2.10191

18. Government Digital Service. *Data Ethics Framework*. UK Government (2020). Available online at: https://www.gov.uk/government/publications/data-ethics-framework/data-ethics-framework-2020 (accessed December 19, 2020).

19. Department of Health and Social Care. *Code of Conduct for Data-Driven Health and Care Technology*. (2019). UK Government. Available online at: https://www.gov.uk/government/publications/code-of-conduct-for-data-driven-health-and-care-technology/initial-code-of-conduct-for-data-driven-health-and-care-technology (accessed December 19, 2020).

20. Jones KH, Ford EM, Lea N, Griffiths LJ, Hassan L, Heys S, et al. Toward the development of data governance standards for using clinical free-text data in health research: position paper. *J Med Internet Res.* (2020) 22:16. doi: 10.2196/16760

21. Okoli C, Pawlowski SD. The Delphi method as a research tool: an example, design considerations and applications. *Inform Manage.* (2004) 42:15–29. doi: 10.1016/j.im.2003.11.002

22. De Villiers MR, De Villiers PJT, Kent AP. The Delphi technique in health sciences education research. *Med Teacher.* (2005) 27:639–43. doi: 10.1080/13611260500069947

23. Bisson JI, Tavakoly B, Witteveen AB, Ajdukovic D, Jehel L, Johansen VJ, et al. TENTS guidelines: development of post-disaster psychosocial care guidelines through a Delphi process. *Br J Psychiatry.* (2010) 196:69–74. doi: 10.1192/bjp.bp.109.066266

24. Jorm AF. Using the Delphi expert consensus method in mental health research. *Aust N Zeal J Psychiatry.* (2015) 49:887–97. doi: 10.1177/0004867415600891

25. Kelly CM, Jorm AF, Kitchener BA, Langlands RL. Development of mental health first aid guidelines for suicidal ideation and behaviour: a Delphi study. *BMC Psychiatry.* (2008) 8:10. doi: 10.1186/1471-244X-8-10

26. Langlands RL, Jorm AF, Kelly CM, Kitchener BA. First aid recommendations for psychosis: using the delphi method to gain consensus between mental health consumers, carers, and clinicians. *Schizophr Bull.* (2008) 34:435–43. doi: 10.1093/schbul/sbm099

27. Murphy C, Thorpe L, Trefusis H, Kousoulis A. Unlocking the potential for digital mental health technologies in the UK: a Delphi exercise. *Bjpsych Open.* (2020) 6:5. doi: 10.1192/bjo.2019.95

28. Sinclair S, Jaggi P, Hack TF, Russell L, McClement SE, Cuthbertson L, et al. Initial validation of a patient-reported measure of compassion: determining the content validity and clinical sensibility among patients living with a life-limiting and incurable illness. *Patient Patient Center Outcomes Res.* (2020) 13:327–37. doi: 10.1007/s40271-020-00409-8

29. McNall M, Foster-Fishman PG. Methods of rapid evaluation, assessment, and appraisal. *Am J Eval.* (2007) 28:151–68. doi: 10.1177/1098214007300895

30. Buckley BS, Murphy AW, MacFarlane AE. Public attitudes to the use in research of personal health information from general practitioners' records: a survey of the Irish general public. *J Med Ethics.* (2011) 37:50–5. doi: 10.1136/jme.2010.037903

31. Jung Y, Choi H, Shim H. Individual willingness to share personal health information with secondary information users in South Korea. *Health Commun.* (2020) 35:659–66. doi: 10.1080/10410236.2019.1582311

32. Iveson MH, Deary IJ. Navigating the landscape of non-health administrative data in Scotland: a researcher's narrative [version 2; peer review: 2 approved]. *Wellcome Open Res.* (2019) 4:97. doi: 10.12688/wellcomeopenres.15336.2

33. Budin-Ljosne I, Teare HJA, Kaye J, Beck S, Bentzen HB, Caenazzo L, et al. Dynamic consent: a potential solution to some of the challenges of modern biomedical research. *BMC Med Ethics.* (2017) 18:10. doi: 10.1186/s12910-016-0162-9

34. Vayena E, Blasimme A. Biomedical big data: new models of control over access, use and governance. *J Bioethical Inquiry.* (2017) 14:501–13. doi: 10.1007/s11673-017-9809-6

35. Jones KH, Laurie G, Stevens L, Dobbs C, Ford DV, Lea N. The other side of the coin: harm due to the non-use of health-related data. *Int J Med Inform.* (2017) 97:43–51. doi: 10.1016/j.ijmedinf.2016.09.010

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.